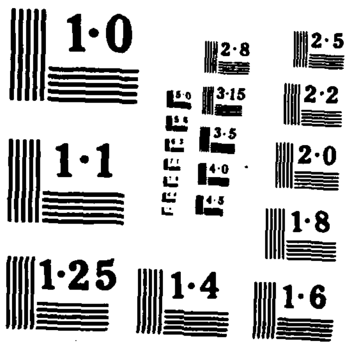
AD-A169 935 SMOOTH NONPARAMETRIC QUANTILE ESTIMATION UNDER CENSORING: SIMULATIONS AND..(U) SOUTH CAROLINA UNIV COLUMBIA DEPT OF STATISTICS W J PADGETT ET AL. MAY 86 1/1

UNCLASSIFIED TR-116 AFOSR-TR-86-0431 AFOSR-84-0156 F/G 12/1 NL


1·0
2·8
2·5
3·15
2·2
3·5
1·1
4·0
2·0
4·5
1·8
1·25
1·4
1·6

DTIC
ELECTE
JUL 2 4 1986
S D
D

(2)

AD-A169 935

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Stat. Tech. Rep. No. 116 (62005-17)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 86-0431
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics	6b. OFFICE SYMBOL (If applicable)
7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research	
6c. ADDRESS (City, State and ZIP Code) University of South Carolina Columbia, SC 29208	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB, DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR, ARO	8b. OFFICE SYMBOL (If applicable) NM
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-84-0156	
8c. ADDRESS (City, State and ZIP Code) Bolling AFB, DC 20332	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A5	

11. TITLE (Include Security Classification)
Smooth Nonparametric Quantile Estimation under Censoring: Simulations and Bootstrap Methods.

12. PERSONAL AUTHOR(S)
W. J. Padgett and L. A. Thombs

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	May, 1986	24

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Nonparametric quantile estimation; Right-censoring; Percentile interval; Bootstrap; Bandwidth selection; Monte Carlo simulations.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Based on right-censored data from a lifetime distribution F_o, a smooth nonparametric estimator of the quantile function $Q^o(p)$ is given by $Q_n(p) = h^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h)dt$, where $\hat{Q}_n(p)$ denotes the product-limit quantile function. Extensive Monte Carlo simulations indicate that at fixed p this kernel-type quantile estimator has smaller mean squared error than $\hat{Q}_n(p)$ for a range of values of the bandwidth h. A method of selecting an "optimal" bandwidth (in the sense of small estimated mean squared error) based on the bootstrap is investigated yielding results consistent with the simulation study. The bootstrap is also used to obtain interval estimates for $Q^o(p)$ after determining the optimal value of h.

DTIC FILE COPY

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Maj. Brian W. Woodruff	(202) 767-5027	NM

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.   UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 86-0431

SMOOTH NONPARAMETRIC QUANTILE ESTIMATION UNDER CENSORING: SIMULATIONS AND BOOTSTRAP METHODS

by

W. J. Padgett* and L. A. Thombs

University of South Carolina
Statistics Technical Report No. 116
62N05-17

DEPARTMENT OF STATISTICS

The University of South Carolina
Columbia, South Carolina 29208

SMOOTH NONPARAMETRIC QUANTILE ESTIMATION UNDER CENSORING: SIMULATIONS AND BOOTSTRAP METHODS

by

W. J. Padgett* and L. A. Thombs

University of South Carolina

Statistics Technical Report No. 116
62N05-17

May, 1986

Department of Statistics
University of South Carolina
Columbia, SC 29208

*Research supported by the U. S. Air Force Office of Scientific Research under grant number AFOSR-84-0156 and. U. S. Army Research Office under grant number MIPR ARO 139-85.

SMOOTH NONPARAMETRIC QUANTILE ESTIMATION UNDER CENSORING: SIMULATIONS AND BOOTSTRAP METHODS

W. J. Padgett and L. A. Thombs

Department of Statistics
University of South Carolina
Columbia, South Carolina 29208

ABSTRACT

Based on right-censored data from a lifetime distribution F_o, a smooth nonparametric estimator of the quantile function $Q^o(p)$ is given by $Q_n(p)=h^{-1}\int_0^1\hat{Q}_n(t)K((t-p)/h)dt$, where $\hat{Q}_n(p)$ denotes the product-limit quantile function. Extensive Monte Carlo simulations indicate that at fixed p this kernel-type quantile estimator has smaller mean squared error than $\hat{Q}_n(p)$ for a range of values of the bandwidth h. A method of selecting an "optimal" bandwidth (in the sense of small estimated mean squared error) based on the bootstrap is investigated yielding results consistent with the simulation study. The bootstrap is also used to obtain interval estimates for $Q^o(p)$ after determining the optimal value of h.

1. INTRODUCTION

Arbitrarily right-censored data arise naturally in industrial life testing and medical studies. In these situations it is important to be able to obtain good nonparametric estimates of various characteristics of the unknown lifetime distribution. One characteristic of the lifetime distribution that is of interest is the *quantile function*. For right-censored data, Sander (1975) proposed estimation of the quantile function by the product-limit quantile estimator, defined by $\hat{Q}_n \equiv \hat{F}_n^{-1}$, where $\hat{F}_n$ denotes the product-limit estimator of the lifetime distribution function F_o (Kaplan and Meier, 1958; Efron, 1967). Sander (1975) and Cheng (1984) obtained some asymptotic properties of $\hat{Q}_n$, and Csörgő (1983) discussed strong approximation results.

The product-limit quantile estimator is a step function with jumps corresponding to the uncensored observations. A smoothed nonparametric estimator of the quantile function from right-censored observations based on the kernel method was proposed by Padgett (1986), extending the complete sample results of Yang (1985). Lio, Padgett, and Yu (1986) and Lio and Padgett (1985) studied some of the asymptotic properties of this kernel estimator, including asymptotic normality and mean square convergence.

In general, the effective performance of nonparametric function estimators is critically dependent on the choice of a "smoothing parameter." If not enough smoothing is done, the estimate will be "rough," showing features which do not represent the function being estimated. On the other hand, if too much smoothing is done, important features of the curve may not show up since they are essentially smoothed away (Marron, 1986). For kernel-type estimators, the smoothing parameter is often called the "bandwidth," and an important question that arises is: Given a set of data, what value(s) of the bandwidth are best to use in calculating the smooth estimator in the sense of minimum mean squared error, or with respect to some other criterion?

The objectives of this paper are two-fold. One is to report results of extensive Monte Carlo simulations which demonstrate the behavior of the mean squared error of the kernel estimator with respect to bandwidth. These simulations provide a method of choosing an optimal bandwidth when the form of the lifetime and censoring distributions are known. Also, they compare the kernel-type estimator with the product-limit quantile estimator. Five commonly used parametric lifetime distributions, two censoring mechanisms, and four different kernel functions are considered in this study, which is an extension of the brief simulations for exponential distributions reported by Padgett (1986).

The second objective is to present a nonparametric method for bandwidth selection based on the bootstrap for right-censored data. This data-based procedure uses the bootstrap to estimate mean squared error, and is both an extension and modification of the methods proposed by Padgett (1986). Bandwidth selection using the bootstrap is important for small and moderately large samples since no exact expressions exist for the mean squared error of the kernel-type quantile estimator. Lio and Padgett (1985) obtained an upper bound on the mean square convergence rate of the kernel quantile estimator under random right-censorship, but the bound is not sharp and does not readily lead to an optimal choice of the bandwidth in the sense of minimum mean squared error. The bootstrap also provides a method of obtaining confidence intervals for the unknown quantiles or more generally, confidence bands for the quantile function. Two such intervals are presented in this paper.

In Section 2, the estimators and some of their asymptotic properties are discussed. The simulation results are reported in Section 3. The bootstrap bandwidth selection procedure is presented in Section 4, and some examples are given which indicate that the data-based nonparametric method yields optimal bandwidths which are consistent with the Monte Carlo results of Section 3. In Section 5, two bootstrap confidence interval procedures are presented along with some convergence results which provide asymptotic validity for the bootstrap in this setting of quantile estimation.

2. NOTATION AND PRELIMINARIES

Let $X_1^o,\ldots,X_n^o$ denote the true survival times of n items or individuals that are censored on the right by a sequence $U_1,\ldots,U_n$, which in general may be either constants or random variables. The X_i^o's are nonnegative, independent, identically distributed random variables with common unknown distribution function F_o and unknown quantile function $Q^o(p) \equiv F_o^{-1}(p) =$

$=\inf\{t:F_o(t)\geq p\}$, $0\leq p\leq 1$.

The observed right-censored data are denoted by the pairs (X_i,Δ_i), $i=1,\ldots,n$, where

$$X_i = \min\{X_i^o,\ U_i\}, \qquad \Delta_i = \begin{cases} 1 \text{ if } X_i^o \leq U_i \\ 0 \text{ if } X_i^o > U_i \end{cases}.$$

Thus, it is known which observations are times of failure or death and which ones are censored or loss times. The nature of the censoring depends on the U_i's. (i) If $U_1,\ldots,U_n$ are fixed constants, the observations are time-truncated. If all U_i's are equal to the same constant, then the case of Type I censoring results. (ii) If all $U_i = X_{(r)}^o$, the rth order statistic of $X_1^o,\ldots,X_n^o$, then the situation is that of Type II censoring. (iii) If $U_1,\ldots,U_n$ constitute a random sample from a distribution H (usually unknown) and are independent of $X_1^o,\ldots,X_n^o$, then (X_i,Δ_i), $i=1,2,\ldots,n$, is called a randomly right-censored sample.

For the asymptotic results of Padgett (1986), Lio, Padgett, and Yu (1986), and Lio and Padgett (1985), the random censorship model (iii) was assumed. For this model the distribution function of each X_i is $F=1-(1-F_o)(1-H)$.

A popular estimator of the survival function $1-F_o(t)$ from the censored sample (X_i,Δ_i), $i=1,\ldots,n$, is the product-limit (PL) estimator of Kaplan and Meier (1958). The PL estimator, which was shown to be "self-consistent" by Efron (1967), is defined as follows. Let (Z_i,Λ_i), $i=1,\ldots,n$, denote the ordered X_i's along with their corresponding Δ_i's. Values of the censored sample will be denoted by the corresponding lower case letters, (x_i,δ_i) and (z_i,λ_i), for the unordered and ordered sample, respectively. Then the PL estimator of $1-F_o(t)$ is

$$\hat{P}_n(t)=\begin{cases} 1, & 0\leq t\leq Z_1, \\ \prod_{i=1}^{k-1}\left(\dfrac{n-i}{n-i+1}\right)^{\Lambda_i}, & Z_{k-1}<t\leq Z_k,\ k=2,\ldots,n \\ 0, & Z_n<t. \end{cases}$$

The PL estimator of $F_o(t)$ is denoted by $\hat{F}_n(t)=1-\hat{P}_n(t)$, and the size of the jump of $\hat{P}_n$ (or $\hat{F}_n$) at Z_j is denoted by s_j. Note that

$s_j=0$ if and only if Z_j is censored for $j<n$, i.e. if and only if $\lambda_j=0$. Define $S_i = \sum_{j=1}^{i} s_j = \hat{F}_n(Z_{i+1})$, $i=1,\ldots,n-1$, and $S_n\equiv 1$.

A natural estimator of $Q^o(p)$ is the PL quantile function $\hat{Q}_n(p)=\inf\{t:\hat{F}_n(t)\geq p\}$ (see, for example, Sander (1975), Cheng (1984), and Csorgo (1983) for some of the properties of $\hat{Q}_n$). Since $\hat{Q}_n$ is a step function with jumps corresponding to the uncensored observations, it is desirable to obtain a smoothed estimator of Q^o. The kernel smoothed $\hat{Q}_n$, considered by Padgett (1986), Lio, Padgett and Yu (1986), and Lio and Padgett (1985), is such an estimator, and is defined as follows: Let $\{h\equiv h_n\}$ be a "bandwidth" sequence of positive numbers so that $h_n\to 0$ as $n\to\infty$, and let K be a bounded probability density function which is zero outside a finite interval $(-c,c)$ and is symmetric about zero. (For asymptotic results, other conditions on h_n, K, and F_o are needed, but these are the only assumptions that will be made here.) Then for $0\leq p\leq 1$, the kernel quantile function estimator is given by

$$
\begin{aligned}
Q_n(p) &= h^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h)dt \\
&= h^{-1}\sum_{i=1}^{n} Z_i \int_{S_{i-1}}^{S_i} K((t-p)/h)dt, \qquad (2.1)
\end{aligned}
$$

where $S_0\equiv 0$. An approximation to $Q_n(p)$ was given by Padgett (1986) as

$$
\tilde{Q}_n(p)=h^{-1}\sum_{i=1}^{n} Z_i s_i K((S_i-p)/h). \qquad (2.2)
$$

Although neither estimator is difficult to compute, (2.2) will be simpler for some kernel functions.

The asymptotic normality, asymptotic equivalence, and mean square convergence of (2.1) and (2.2) were studied by Lio, Padgett and Yu (1986) and Lio and Padgett (1985). However, no small sample properties have been derived. In fact, due to the mathematical complications introduced by censoring, an exact expression for the mean squared error of $Q_n(p)$ for small n is not available. Thus, a bandwidth value minimizing the exact mean squared error of $Q_n(p)$ cannot be obtained. A practical method for selecting bandwidth is discussed in Section 4. First, in the next section, a large simulation study is reported which gives comparisons of Q_n and $\tilde{Q}_n$ with $\hat{Q}_n$ and gives an indication of the behavior of these estimators with respect to the bandwidth values, the censoring mechanism, the kernel function, and sample size, using the mean squared error criterion.

3. COMPARISON OF ESTIMATORS: SIMULATION RESULTS

A Monte Carlo simulation was performed for five families of lifetime distributions that are commonly used in life testing. These distributions are shown in Table 1. Two censoring distributions H were used: exponential with density $h(u) = \lambda e^{-\lambda u}$, $u>0$, $\lambda>0$, and uniform on the interval $(0,\lambda)$, $\lambda>0$. In addition, three different kernel functions were chosen as $K_1(x)=1-|x|$, $|x|\leq 1$ (triangular), $K_2(x)=3/4(1-x^2)$, $|x|\leq 1$ (quadratic), and $K_3(x)=1$, $|x|\leq 0.5$ (uniform). Also, the uniform kernel on [-1,1] was used, producing similar results as K_3.

TABLE 1. Lifetime Distributions Used in Simulations

Distribution	Density	Notation
Exponential	$f(x)=\beta\exp(-\beta x), x>0$	$E(\beta): \beta=1$
Weibull	$f(x)=\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right],\ x>0$	$W(\alpha,\beta)$: $(\alpha,\beta)=(0.5,1),(2,1),(2,5)$
Gamma	$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}\exp(-x/\beta),\ x>0$	$G(\alpha,\beta)$: $(\alpha,\beta)=(0.5,1),(2,1),(2,5)$
Lognormal	$f(x)=\frac{1}{(2\pi)^{1/2}\beta x}\exp\left[-\frac{(\log x-\alpha)^2}{2\beta^2}\right],\ x>0$	$L(\alpha,\beta)$: $(\alpha,\beta)=(0,1),(2,0.5)$
Inverse Gaussian	$f(x)=\left(\frac{\lambda}{2\pi x^3}\right)^{1/2}\exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right],\ x>0$	$IG(\mu,\lambda)$: $(\mu,\lambda)=(1,0.25),(3,1)$

The parameter λ of the censoring distribution was determined to give either 30% or 50% censoring. That is, λ was determined so that the probability of a censored observation, $Pr(X^o>U)=0.3$ or 0.5, at least approximately. This probability was calculated by numerical integration using the midpoint rule when it could not be obtained exactly. The value of λ is reported in the resulting table for each case.

Bandwidth values of h=0.01 (.02) 0.61 were used for quantiles at p = 0.10, 0.25, 0.50, 0.75, 0.90, and 0.95. Sample sizes of n=20, 50, 100, and 300 were chosen, although only the results for n=50 and 100 are shown in the tables presented here for brevity.

In each case simulated (i.e. each distribution, kernel, bandwidth, p, and sample size combination), 1000 censored samples were generated using the random number generators in the International Mathematical and Statistical Library (IMSL, 1982) on an IBM 370 computer. In particular, for uniform random number generation, the IMSL subroutine GGUBS was used; for exponential random numbers, GGEXN; for Weibull, GGWIB; for gamma, GGAMR; and for lognormal, GGNLG. For generating random numbers from the inverse Gaussian distribution, the method discussed in Michael, Schucany, and Haas (1976) was utilized. From the 1000 samples, the estimated mean squared errors (MSE) of the estimators $Q_n(p)$, $\tilde{Q}_n(p)$ and $\hat{Q}_n(p)$ were computed, and the ratios of these estimated mean squared errors, A=(MSE $\hat{Q}_n$/MSE Q_n) and B=(MSE $\hat{Q}_n$/MSE $\tilde{Q}_n$), were calculated.

Some of the results of the simulations are shown in Tables 2-9. In each case, for each p, there is a range of bandwidth values for which $Q_n(p)$ has smaller estimated mean squared error than that of $\hat{Q}_n(p)$. The same behavior of $\tilde{Q}_n$ was observed, except that the best bandwidth values were generally larger than those for Q_n, indicating that more smoothing is required for $\tilde{Q}_n$.

Parzen (1979) indicated that kernel estimators of quantile functions do not generally give good estimates for p near zero or one since quantile functions are usually nonintegrable. This is quite noticeable for Q_n and $\tilde{Q}_n$ in Tables 2-9 for p near one, although for some values of h, Q_n is still better than the PL quantile estimator. Also, it should be noted that as $h\to 0$ for fixed n, $Q_n(p)\to\hat{Q}_n(p)$ and hence the ratios of mean squared errors

TABLE 2. RATIOS OF MEAN SQUARE ERRORS WITH TRIANGULAR KERNEL

LIFE DISTRIBUTION: E(1), CENSORING DISTRIBUTION: E(3/7)

n = 100 (30% CENSORING)

p		h: 0.01	0.03	0.05	0.07	0.09	0.11	0.13	0.15	0.19	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.02	1.02	1.07	1.12	1.17	1.23	1.28	1.32	1.33	1.30	1.18	0.93	0.76	0.55	0.44	0.32	0.26	0.19
	B	0.38	1.24	1.30	1.38	1.46	1.55	1.63	1.71	1.80	1.80	1.70	1.36	1.11	0.78	0.61	0.43	0.34	0.25
0.25	A	1.03	1.06	1.09	1.11	1.13	1.16	1.18	1.20	1.24	1.26	1.28	1.29	1.25	1.13	1.02	0.82	0.70	0.53
	B	0.15	1.19	1.24	1.27	1.30	1.33	1.37	1.40	1.47	1.50	1.56	1.64	1.64	1.57	1.46	1.22	1.05	0.80
0.50	A	1.02	1.05	1.07	1.09	1.11	1.12	1.14	1.15	1.16	1.16	1.15	1.10	1.05	0.92	0.80	0.62	0.57	0.57
	B	0.05	1.09	1.24	1.29	1.32	1.34	1.36	1.38	1.42	1.44	1.47	1.50	1.52	1.49	1.44	1.31	1.10	0.98
0.75	A	1.02	1.02	1.02	1.03	1.04	1.04	1.03	1.01	0.97	0.94	0.87	0.93	1.09	1.34	1.41	1.32	1.16	0.92
	B	0.04	0.50	1.02	1.20	1.31	1.36	1.40	1.44	1.53	1.52	1.48	1.69	1.70	1.66	1.54	1.27	1.08	0.85
0.90	A	1.03	1.11	1.18	1.24	1.30	1.38	1.56	1.73	1.77	1.64	1.32	0.94	0.78	0.62	0.54	0.47	0.43	0.38
	B	0.05	0.28	0.68	0.81	0.82	1.17	1.73	1.92	1.76	1.59	1.27	0.92	0.77	0.61	0.54	0.47	0.43	0.38
0.95	A	1.05	1.16	1.24	1.32	1.23	1.05	0.88	0.74	0.57	0.51	0.43	0.36	0.32	0.29	0.27	0.25	0.24	0.22
	B	0.06	0.22	0.26	0.87	1.08	1.04	0.93	0.81	0.63	0.56	0.47	0.38	0.34	0.30	0.28	0.26	0.24	0.23

n = 100 (30% CENSORING)

p		h: 0.01	0.03	0.05	0.07	0.09	0.11	0.13	0.15	0.19	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.02	1.08	1.15	1.21	1.27	1.33	1.38	1.41	1.37	1.30	1.10	0.75	0.57	0.37	0.28	0.19	0.15	0.11
	B	1.13	1.20	1.28	1.35	1.43	1.51	1.59	1.65	1.68	1.62	1.40	0.96	0.71	0.45	0.34	0.23	0.18	0.13
0.25	A	1.02	1.05	1.08	1.10	1.13	1.16	1.18	1.21	1.25	1.27	1.29	1.27	1.21	1.04	0.89	0.66	0.53	0.37
	B	0.96	1.10	1.15	1.17	1.21	1.24	1.27	1.31	1.37	1.40	1.45	1.48	1.45	1.29	1.12	0.85	0.68	0.47
0.50	A	1.01	1.03	1.05	1.07	1.09	1.11	1.12	1.13	1.15	1.15	1.14	1.07	0.98	0.80	0.66	0.45	0.38	0.37
	B	0.42	1.07	1.10	1.13	1.16	1.18	1.20	1.22	1.26	1.28	1.31	1.30	1.26	1.12	0.99	0.75	0.59	0.53
0.75	A	1.03	1.05	1.07	1.08	1.09	1.09	1.09	1.07	1.00	0.93	0.78	0.79	0.98	1.33	1.37	1.07	0.83	0.56
	B	0.08	1.07	1.21	1.27	1.32	1.36	1.40	1.44	1.49	1.50	1.48	1.37	1.47	1.56	1.40	0.98	0.75	0.51
0.90	A	1.09	1.23	1.31	1.35	1.37	1.43	1.67	1.95	1.87	1.58	1.07	0.66	0.51	0.39	0.33	0.28	0.25	0.22
	B	0.06	0.58	1.09	1.28	1.28	1.66	2.12	2.18	1.72	1.43	0.99	0.63	0.50	0.38	0.33	0.28	0.25	0.22
0.95	A	1.05	1.17	1.30	1.55	1.50	1.18	0.90	0.71	0.49	0.43	0.35	0.28	0.25	0.22	0.20	0.18	0.17	0.16
	B	0.06	0.34	0.41	1.21	1.33	1.13	0.90	0.73	0.51	0.45	0.36	0.28	0.25	0.22	0.20	0.19	0.18	0.16

A = (MSE $\hat{Q}_n$)/(M[illegible] Q_n), B = (MSE $\hat{Q}_n$)/(MSE $\bar{Q}_n$)

TABLE 3. RATIOS OF MEAN SQUARE ERRORS WITH QUADRATIC KERNEL

LIFE DISRIBUTION: E(1) , CENSORING DISTRIBUTION: E(3/7)
n = 100 (30% CENSORING)

p		h: 0.01	0.05	0.11	0.15	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.02	1.16	1.36	1.44	1.23	0.97	0.61	0.44	0.28	0.21	0.14	0.11	0.08
	B	0.79	1.29	1.55	1.70	1.56	1.24	0.77	0.54	0.33	0.25	0.16	0.13	0.09
0.25	A	1.02	1.08	1.17	1.23	1.29	1.30	1.25	1.15	0.92	0.75	0.52	0.40	0.27
	B	0.82	1.16	1.26	1.33	1.43	1.48	1.48	1.40	1.16	0.95	0.66	0.51	0.34
0.50	A	1.01	1.06	1.12	1.14	1.15	1.12	1.01	0.90	0.68	0.53	0.33	0.29	0.30
	B	0.77	1.11	1.19	1.24	1.30	1.31	1.28	1.20	1.00	0.84	0.59	0.44	0.41
0.75	A	1.03	1.07	1.09	1.05	0.86	0.67	0.73	1.01	1.37	1.20	0.75	0.54	0.36
	B	0.12	1.21	1.38	1.45	1.49	1.42	1.24	1.41	1.46	1.16	0.70	0.51	0.34
0.90	A	1.10	1.32	1.44	2.06	1.22	0.77	0.48	0.38	0.30	0.26	0.22	0.20	0.18
	B	0.07	1.18	1.83	2.28	1.21	0.78	0.48	0.39	0.30	0.26	0.22	0.21	0.18
0.95	A	1.06	1.34	0.96	0.55	0.34	0.28	0.23	0.21	0.19	0.17	0.16	0.15	0.14
	B	0.07	0.40	1.07	0.61	0.37	0.30	0.24	0.22	0.19	0.18	0.16	0.16	0.15

A = (MSE $\hat{Q}_n$)/(MSE Q_n), B = (MSE $\hat{Q}_n$)/(MSE $\tilde{Q}_n$)

TABLE 4. RATIOS OF MEAN SQUARE ERRORS WITH UNIFORM KERNEL

LIFE DISTRIBUTION: E(1) , CENSORING DISTRIBUTION: E(3/7)
n = 100 (30% CENSORING)

p		h: 0.01	0.05	0.11	0.15	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.02	1.10	1.24	1.32	1.45	1.48	1.34	1.16	0.87	0.71	0.51	0.41	0.30
	B	0.64	1.25	1.42	1.47	1.67	1.71	1.63	1.46	1.09	0.88	0.63	0.49	0.36
0.25	A	1.02	1.06	1.12	1.16	1.21	1.24	1.29	1.30	1.30	1.28	1.23	1.17	1.03
		0.24	0.93	1.11	1.19	1.27	1.32	1.41	1.45	1.49	1.47	1.46	1.41	1.29
0.50	A	1.00	1.03	1.08	1.11	1.13	1.14	1.15	1.14	1.10	1.06	0.97	0.89	0.76
	B	0.06	0.68	1.02	1.10	1.16	1.19	1.25	1.25	1.30	1.27	1.21	1.15	1.06
0.75	A	1.02	1.05	1.08	1.08	1.07	1.02	0.90	0.78	0.60	0.48	0.41	0.68	1.22
	B	0.03	0.52	1.03	1.18	[illegible]	1.26	1.24	1.23	1.11	0.96	0.39	0.66	1.20
0.90	A	1.07	1.26	1.35	1.35	1.61	2.06	1.18	0.82	0.55	0.45	0.36	0.31	0.27
	B	0.03	0.47	1.10	1.07	1.34	2.12	1.32	0.90	0.58	0.47	0.37	0.32	0.28
0.95	A	1.03	1.20	1.63	1.08	0.55	0.42	0.32	0.28	0.24	0.22	0.20	0.19	0.18
	B	0.03	0.36	1.06	1.30	0.66	0.47	0.35	0.30	0.25	0.23	0.21	0.20	0.18

A = (MSE $\hat{Q}_n$)/(MSE Q_n), B = (MSE $\hat{Q}_n$)/(MSE $\tilde{Q}_n$)

TABLE 5. RATIOS OF MEAN SQUARE ERRORS WITH TRIANGULAR KERNEL

LIFE DISTRIBUTION: E(1) , CENSORING DISTRIBUTION: U(0, 3.1941)
n = 100 (APPROX. 30% CENSORING)

p		h 0.01	0.05	0.11	0.15	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.02	1.15	1.33	1.43	1.33	1.13	0.77	0.58	0.38	0.29	0.20	0.16	0.11
	B	1.10	1.27	1.50	1.65	1.65	1.43	0.98	0.73	0.46	0.34	0.23	0.18	0.13
0.25	A	1.02	1.07	1.16	1.21	1.28	1.31	1.29	1.23	1.04	0.89	0.65	0.52	0.36
	B	1.04	1.14	1.23	1.30	1.40	1.46	1.49	1.46	1.29	1.11	0.83	0.66	0.46
0.50	A	1.01	1.05	1.12	1.15	1.16	1.15	1.07	0.98	0.79	0.65	0.48	0.43	0.45
	B	0.48	1.09	1.18	1.24	1.30	1.32	1.32	1.28	1.20	1.15	1.07	0.86	0.80
0.75	A	1.02	1.08	1.11	1.10	1.03	0.97	1.17	1.52	1.81	1.52	0.96	0.70	0.46
	B	0.08	1.23	1.46	1.59	1.66	1.43	2.24	2.20	1.64	1.20	0.75	0.57	0.40
0.90	A	1.07	1.35	1.90	1.76	0.73	0.46	0.29	0.23	0.18	0.16	0.13	0.12	0.11
	B	0.03	0.21	0.30	0.60	0.47	0.36	0.25	0.21	0.17	0.15	0.13	0.12	0.11
0.95	A	1.02	1.02	0.28	0.17	0.11	0.09	0.08	0.07	0.06	0.06	0.06	0.05	0.05
	B	0.02	0.05	0.21	0.16	0.11	0.09	0.08	0.07	0.06	0.06	0.06	0.05	0.05

A = (MSE $\hat{Q}_n$)/(MSE Q_n), B = (MSE $\hat{Q}_n$)/MSE $\tilde{Q}_n$)

TABLE 6. RATIOS OF MEAN SQUARE ERRORS WITH TRIANGULAR KERNEL

LIFE DISTRIBUTION: W(2,1) , CENSORING DISTRIBUTION: E(0.425)
n = 100 (APPROX. 30% CENSORING)

p		h 0.01	0.05	0.11	0.15	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.02	1.14	1.34	1.41	1.39	1.42	1.56	1.74	2.13	2.48	3.02	3.27	3.16
	B	0.82	1.13	1.23	1.16	1.07	1.08	1.20	1.34	1.67	1.99	2.63	3.10	3.56
0.25	A	1.02	1.08	1.16	1.23	1.34	1.41	1.46	1.45	1.42	1.42	1.47	1.54	1.71
	B	0.47	1.09	1.17	1.22	1.29	1.32	1.28	1.23	1.17	1.15	1.17	1.22	1.36
0.50	A	1.03	1.07	1.14	1.18	1.24	1.28	1.33	1.36	1.39	1.39	1.36	1.45	1.73
	B	0.11	1.06	1.13	1.18	1.25	1.30	1.37	1.42	1.50	1.55	1.60	1.66	1.73
0.75	A	1.03	1.07	1.13	1.16	1.17	1.10	1.33	1.40	0.77	0.46	0.25	0.18	0.12
	B	0.04	1.05	1.15	1.22	1.32	1.36	1.44	1.31	0.70	0.44	0.24	0.17	0.12
0.90	A	1.04	1.14	1.20	1.18	0.33	0.19	0.11	0.08	0.06	0.06	0.05	0.04	0.04
	B	0.02	1.02	1.25	1.09	0.35	0.20	0.11	0.09	0.07	0.06	0.05	0.04	0.04
0.95	A	1.07	1.23	0.38	0.17	0.10	0.08	0.06	0.05	0.05	0.04	0.04	0.04	0.04
	B	0.02	0.54	0.45	0.20	0.11	0.08	0.06	0.05	0.05	0.05	0.04	0.04	0.04

A = (MSE $\hat{Q}_n$)/(MSE Q_n), B = (MSE $\hat{Q}_n$)/(MSE $\tilde{Q}_n$)

TABLE 7. RATIOS OF MEAN SQUARE ERRORS WITH TRIANGULAR KERNEL

LIFE DISTRIBUTION: G(2,1) , CENSORING DISTRIBUTION: E(0.415)
n = 100 (APPROX. 50% CENSORING)

p		h 0.01	0.05	0.11	0.15	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.04	1.15	1.39	1.58	1.80	1.96	2.23	2.37	2.36	2.15	1.66	1.33	0.92
	B	0.78	1.23	1.41	1.46	1.55	1.69	2.03	2.31	2.69	2.75	2.40	1.98	1.37
0.25	A	1.02	1.08	1.17	1.23	1.34	1.41	1.54	1.62	1.72	1.78	1.81	1.78	1.60
	B	0.21	1.16	1.25	1.32	1.42	1.49	1.57	1.62	1.71	1.80	1.95	2.06	2.16
0.50	A	1.01	1.06	1.12	1.16	1.20	1.22	1.21	1.17	1.07	0.97	0.79	0.76	0.88
	B	0.04	1.11	1.24	1.30	1.39	1.44	1.53	1.58	1.65	1.65	1.63	1.56	1.61
0.75	A	1.01	1.07	1.14	1.14	1.08	0.99	1.14	1.40	1.55	1.31	0.87	0.65	0.45
	B	0.02	0.69	1.23	1.33	1.32	1.16	1.75	1.73	1.37	1.06	0.70	0.54	0.39
0.90	A	1.05	1.23	1.47	1.85	1.27	0.86	0.55	0.44	0.34	0.29	0.25	0.23	0.20
	B	0.02	0.31	0.60	1.27	1.04	0.77	0.52	0.43	0.33	0.29	0.25	0.23	0.20
0.95	A	1.06	1.26	0.82	0.50	0.32	0.26	0.21	0.19	0.17	0.16	0.15	0.14	1.13
	B	0.03	0.13	0.82	0.57	0.36	0.29	0.23	0.21	0.18	0.17	0.15	0.14	1.13

$A = (\mathrm{MSE}\ \hat{Q}_n)/(\mathrm{MSE}\ Q_n)$, $B = (\mathrm{MSE}\ \hat{Q}_n)/(\mathrm{MSE}\ \tilde{Q}_n)$

TABLE 8. RATIOS OF MEAN SQUARE ERRORS WITH TRIANGULAR KERNEL

LIFE DISTRIBUTION: L(0,1), CENSORING DISTRIBUTION: E(0.274)
n = 100 (APPROX. 30% CENSORING)

p		h 0.01	0.05	0.11	0.15	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.07	1.19	1.41	1.57	1.74	1.85	2.00	1.98	1.65	1.31	0.85	0.62	0.39
	B	1.11	1.24	1.43	1.48	1.56	1.69	1.97	2.13	2.03	1.70	1.11	0.80	0.49
0.25	A	1.02	1.08	1.16	1.20	1.25	1.27	1.30	1.29	1.20	1.08	0.84	0.68	0.45
	B	0.95	1.14	1.23	1.29	1.36	1.41	1.47	1.49	1.44	1.35	1.10	0.89	0.61
0.50	A	1.03	1.07	1.11	1.11	1.07	1.01	0.88	0.76	0.56	0.42	0.25	0.20	0.18
	B	0.41	1.17	1.23	1.26	1.27	1.24	1.13	1.03	0.81	0.66	0.45	0.32	0.26
0.75	A	1.02	1.06	1.04	0.96	0.69	0.49	0.43	0.50	0.71	0.90	1.14	1.15	0.95
	B	0.10	1.27	1.41	1.46	1.43	1.37	0.90	0.91	1.12	1.28	1.34	1.21	0.92
0.90	A	1.10	1.21	1.28	1.86	2.34	1.90	1.26	0.99	0.74	0.64	0.53	0.48	0.42
	B	0.12	1.79	2.54	3.32	2.44	1.79	1.18	0.94	0.72	0.62	0.52	0.47	0.42
0.95	A	1.07	1.31	1.58	1.10	0.70	0.57	0.46	0.41	0.36	0.34	0.31	0.30	0.28
	B	0.10	0.51	1.51	1.07	0.70	0.58	0.47	0.42	0.37	0.34	0.31	0.30	0.28

$A = (\mathrm{MSE}\ \hat{Q}_n)/(\mathrm{MSE}\ Q_n)$, $B = (\mathrm{MSE}\ \hat{Q}_n)/(\mathrm{MSE}\ \tilde{Q}_n)$

TABLE 9. RATIOS OF MEAN SQUARE ERRORS WITH TRIANGULAR KERNEL

LIFE DISTRIBUTION: IG(3,1) , CENSORING DISTRIBUTION: E(0.182)
n = 100 (APPROX. 30% CENSORING)

p		h 0.01	0.05	0.11	0.15	0.21	0.25	0.31	0.35	0.41	0.45	0.51	0.55	0.61
0.10	A	1.04	1.12	1.28	1.39	1.45	1.49	1.50	1.40	1.05	0.78	0.46	0.32	0.19
	B	1.13	1.20	1.33	1.34	1.34	1.40	1.53	1.55	1.30	1.00	0.59	0.41	0.23
0.25	A	1.01	1.05	1.09	1.10	1.07	1.02	0.93	0.84	0.66	0.53	0.36	0.26	0.16
	B	0.96	1.12	1.18	1.21	1.21	1.18	1.11	1.02	0.83	0.68	0.46	0.34	0.21
0.50	A	1.02	1.04	1.04	1.02	0.92	0.81	0.61	0.47	0.29	0.20	0.11	0.09	0.07
	B	0.71	1.17	1.19	1.19	1.12	1.04	0.84	0.69	0.48	0.37	0.24	0.16	0.11
0.75	A	1.04	1.07	0.94	0.80	0.55	0.41	0.38	0.42	0.57	0.71	0.99	1.19	1.40
	B	0.17	1.58	1.76	1.82	1.82	1.80	1.10	1.01	1.12	1.28	1.55	1.68	1.69
0.90	A	1.07	1.27	1.58	2.22	2.60	2.20	1.57	1.28	1.00	0.88	0.75	0.68	0.61
	B	0.16	1.11	1.73	2.97	2.42	1.92	1.40	1.17	0.94	0.83	0.72	0.66	0.59
0.95	A	1.04	1.17	0.83	0.59	0.42	0.35	0.30	0.27	0.24	0.23	0.21	0.21	0.20
	B	0.10	0.23	0.70	0.55	0.41	0.35	0.30	0.27	0.24	0.23	0.21	0.21	0.20

$A = (\mathrm{MSE}\ \hat{Q}_n)/(\mathrm{MSE}\ Q_n)$, $B = (\mathrm{MSE}\ \hat{Q}_n)/(\mathrm{MSE}\ \tilde{Q}_n)$

should be near one for small h. This is not the case for $\tilde{Q}_n$, as pointed out by Padgett (1986).

With respect to the kernel functions, the results are quite similar for all three. However, the bandwidth value giving the largest ratio of mean squared errors is generally slightly larger for the uniform kernel than for the triangular or quadratic kernels, indicating that more smoothing is needed.

In all cases, the bandwidth value giving the largest ratio of estimated mean squared errors for Q_n tends to increase with p up to about p=0.75, and then decrease for larger p. This indicates that more smoothing is needed in the middle of the distribution than in the tails to decrease the mean squared error of the estimator.

Increasing the amount of censoring from 30% to 50% seems to have little effect on the estimated ratios.

4. BANDWIDTH SELECTION USING THE BOOTSTRAP

The simulation results of Section 3 indicate reasonable ranges of the bandwidth to use in practice, if the forms of the lifetime distribution and censoring mechanism are known. However, in general, the forms of the distributions are not known; this is the reason for using a nonparametric estimator. Since the proposed estimator $Q_n(p)$ is nonparametric, it is desired to use a bandwidth selection method which does not require the parametric assumptions of the results of Section 3. That is, given the right-censored sample of size n, what is an "optimal" bandwidth value to use in calculating $Q_n(p)$? The bootstrap for censored data provides a solution to this problem for a minimum mean squared error optimality criterion.

It is proposed to estimate the mean squared error of $Q_n(p)$, MSE $(Q_n(p))$, as a function of H (for fixed p) by the bootstrap method and to choose the value of h which minimizes the estimated MSE $(Q_n(p))$. Let (x_i, δ_i), i=1,...,n, denote the observed censored sample. A bootstrap replicate (x_i^*, δ_i^*), i=1,...,n, is obtained by randomly drawing with replacement from the set of n bivariate observations (x_i, δ_i). Note that this simple resampling scheme makes no use of the estimated survival distribution or censoring distribution, but has been shown to give the same results as the bootstrap based on a resampling scheme which reflects the censoring structure of the data (see Efron, 1981).

Denote the product-limit estimate of $Q^o(p)$ based on a bootstrap sample by $\hat{Q}_n^*(p)$ and let $Q_n^*(p)$ denote the kernel estimate from bootstrap data; that is,

$$Q_n^*(p) = h^{-1}\int_0^1 \hat{Q}_n^*(t)\, K((t-p)/h)dt.$$

Based on a large number, B, of bootstrap replicates, an estimator of the variance of $Q_n(p)$ is

$$\hat{V}(h) = (B-1)^{-1}\{\sum_{i=1}^{B} [Q_{ni}^*(p)]^2 - [\sum_{i=1}^{B} Q_{ni}^*(p)]^2/B\}, \tag{4.1}$$

and the bootstrap estimate of $\mathrm{Bias}[Q_n(p)]$ is

$$\hat{B}_1(h) = B^{-1}\sum_{i=1}^{B} Q_{ni}^*(p) - Q_n(p).$$

Denote the bootstrap estimate of the mean squared error of $Q_n(p)$, $MSE_{Q_n(p)}$, by $MSE^*_{Q_n(p)}$.

In many of the simulations described in Section 3, as with most kernel-type function estimators, for fixed p, as the bandwidth h increased, the square of the bias of $Q_n(p)$ tended to increase while the variance tended to decrease. Thus, $MSE_{Q_n(p)}(h)$ should be a decreasing and then increasing function of h, and the bootstrap estimate, $MSE^*_{Q_n(p)}(h)$, should yield a value of h giving an approximate minimum value of $MSE_{Q_n(p)}(h)$. However, in many situations encountered in this study, $MSE^*_{Q_n(p)}(h)$ was strictly decreasing in h. This was due to both $Q_n(p)$ and the estimate from a bootstrap sample, $Q_n^*(p)$, being oversmoothed, and hence quite close together. This resulted in a poor estimate of the bias in a "variance-plus-bias^2" estimation of the mean squared error of $Q_n(p)$. Therefore, the bootstrap estimate of bias was modified by using the PL quantile function from (x_i, δ_i), which does not depend on h, instead of $Q_n(p)$. Hence, the bias estimate was modified to

$$\hat{B}_2(h) = B^{-1} \sum_{i=1}^{B} Q_{ni}^{*}(p) - \hat{Q}_n(p).$$

As h gets large (i.e. the kernel estimate is oversmoothed), $[\hat{B}_2(h)]^2$ tends to increase. A bootstrap estimate of $MSE_{Q_n(p)}(h)$ then obtained by

$$MSE^{*}_{Q_n(p)}(h) = \hat{V}(h) + \hat{B}_2^2(h). \quad (4.2)$$

The value of h is chosen to minimize (4.2), yielding an estimated bandwidth $\hat{h}(p)$.

To illustrate the procedure and to give an indication of how well it performs, a random sample of size n=100 was generated from the exponential life distribution E(1) with E(3/7) censoring distribution as in Table 2. The functions $MSE^{*}_{Q_n(p)}(h)$ from B=300 bootstrap samples at each h and p are shown in Figure 1. The triangular kernel function was used in these calculations. The estimated ratios of mean squared errors from Table 2 are shown as functions of h in Figure 2. Table 10 shows the "best" bandwidths from Table 2, $h_R(p)$, and $\hat{h}(p)$ from Figure 1. Note the very close agreement of these bandwidth values. Figure 3 shows the true quantile function Q^o and the kernel estimate $Q_n(p)$ using the estimates of h as 0.17 ($0<p<.2$), 0.23 ($.2\leq p<.5$), 0.25 ($.5\leq p<.7$), 0.47 ($.7\leq p<.8$), 0.11 ($.8\leq p<.95$), and 0.07 ($.95\leq p<1.0$). In general, a value of $\hat{h}(p)$ can be estimated for each value of p for which $Q_n(p)$ is to be plotted.

TABLE 10. Comparison of Best Bandwidths from Figures 1 and 2

p	$h_R(p)$	$\hat{h}(p)$
0.10	0.15	0.17
0.25	0.25	0.23
0.50	0.21	0.25
0.75	0.45	0.47
0.90	0.15	0.11
0.95	0.07	0.07

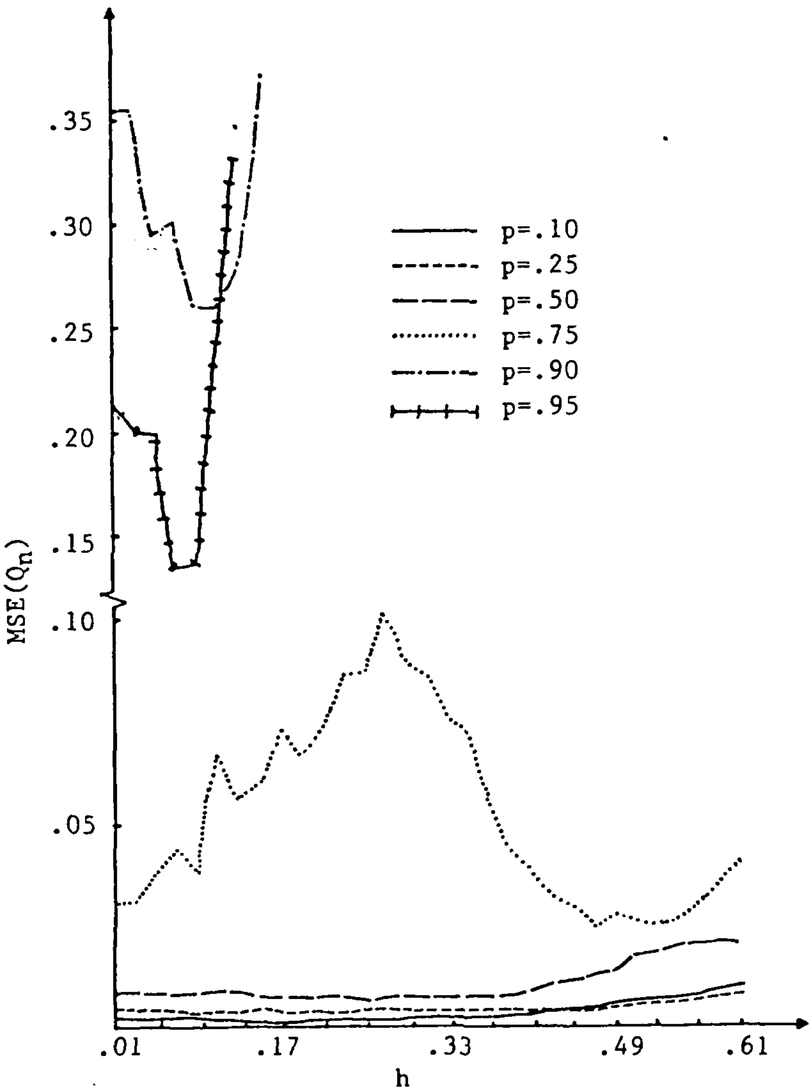


Figure 1. Bootstrap Estimates of MSE(Q_n) vs. h for Simulated Sample

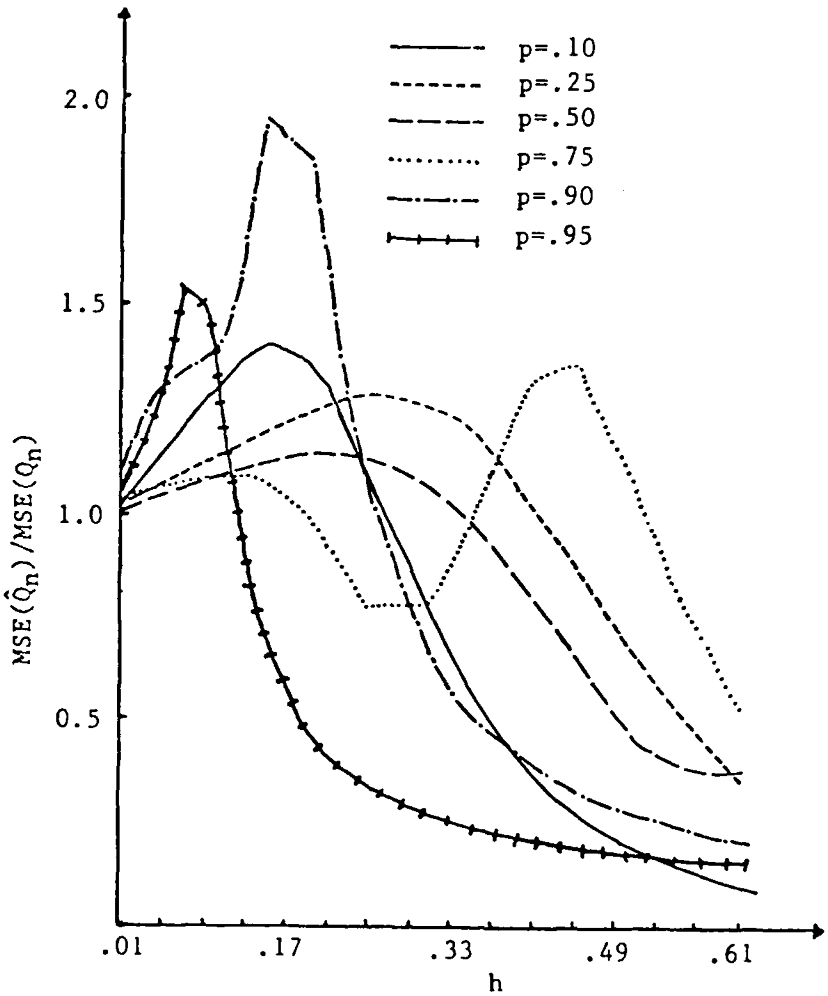


Figure 2. Simulated MSE Ratios

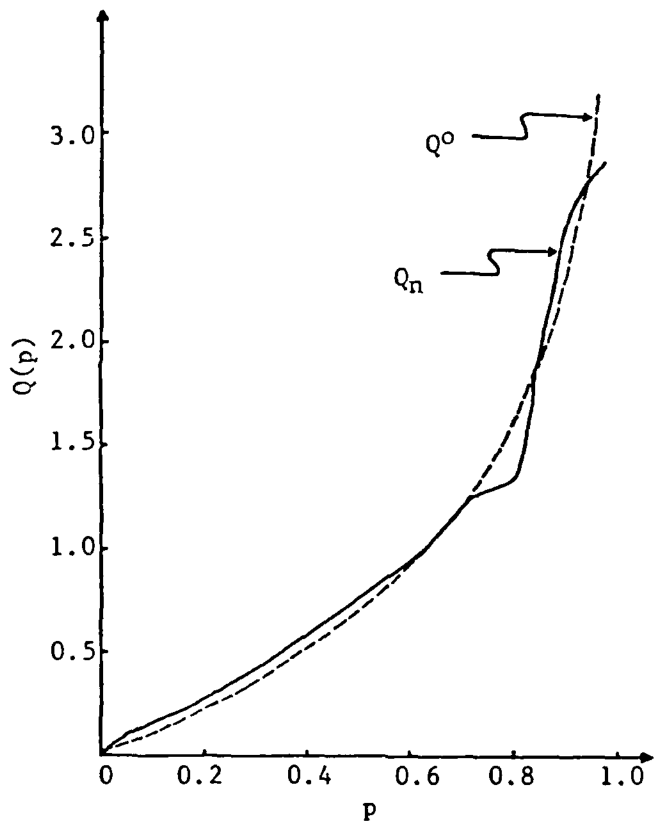


Figure 3. Quantile Estimate from Simulated Exponential Data

5. INTERVAL ESTIMATION

The bootstrap can also be used to obtain interval estimates for $Q^o(p)$. Assume that h has been determined and p is fixed.

The bootstrap estimate of the standard error of $Q_n(p)$ given by equation (4.1) can be used to define an approximate $(1-\alpha)100\%$ confidence interval for $Q^o(p)$,

$$\left(Q_n(p) - z_{1-\alpha/2}\sqrt{\hat{V}(h)},\ Q_n(p) + z_{1-\alpha/2}\sqrt{\hat{V}(h)}\right), \tag{5.1}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ percentile point of the standard normal distribution. This interval requires no additional bootstrap calculations than those involved in selecting h. Efron (1986) has shown that the symmetric interval of the form $(\hat{\theta}\pm\hat{\sigma}z)$ is "correct" if the statistic $\hat{\theta}$ has a normal distribution. Asymptotic normality of $Q_n(p)$ had been established (Lio, Padgett, and Yu, 1986), but for small to moderately large samples and p near 0 or 1 there is some skewness in the distribution of $Q_n(p)$. Furthermore, Efron's results on the validity of the standard bootstrap interval $(\hat{\theta}\pm\hat{\sigma}z)$ refer to the parametric bootstrap in which resampling is from the parametric MLE of the distribution function. Although easily computable, the interval given in (5.1) may be inaccurate, since small sample skewness of the distribution of $Q_n(p)$ will not be reflected. Since the nonparametric bootstrap is used here, an interval which requires no normality (or symmetry) assumptions may be more appropriate in this setting of quantile estimation.

The above discussion suggests a second approach based on the bootstrap percentile interval method. The idea is to use the bootstrap values $Q^*_{ni}(p)$, $i=1,\ldots,B$, to estimate the actual distribution, G, of $Q_n(p)$. Given n and p, the (Monte Carlo estimate of the) bootstrap distribution of $Q_n(p)$ is defined as

$$G^*(k) = P^*(Q_n(p)\le k) = \frac{\#\{Q^*_{ni}(p)\le k\}}{B}. \tag{5.2}$$

While $B = 300$ is sufficient to estimate standard errors, a larger number of bootstrap values $Q^*_{ni}(p)$ is needed to obtain adequate estimates of G. When B is too small, the bootstrap may yield poor estimates of the tail behavior of G. Generally, B=1000 is considered large enough.

The percentile interval uses quantiles of G^* to estimate the quantiles of the true distribution G. An approximate 100(1-α)% confidence interval for $Q^o(p)$ is defined by

$$[G^{*-1}(\alpha),\ G^{*-1}(1-\alpha)]. \tag{5.3}$$

Note that this interval requires additional computations to those involved with bandwidth selection, but these calculations are minimal since the value of h has been determined.

As an illustration of the methods presented in Section 4 and equations (5.1) and (5.3), consider the censored data set of n=100 observations given in Table 11. This data set was generated from the exponential life and censoring distributions used in Table 2 (that is, 30% censoring). The values of $\hat{h}(p)$ were determined from

TABLE 11. Simulated Censored Sample

x_i	δ_i	x_i	δ_i	x_i	δ_i	x_i	δ_i
0.447	0	0.400	1	0.956	0	0.449	0
0.773	1	0.042	1	0.005	1	0.483	1
0.750	1	0.532	1	0.908	1	0.302	0
0.033	1	0.308	1	0.829	0	0.840	1
1.049	1	0.077	0	0.580	0	0.020	0
0.397	1	0.884	1	0.305	1	0.157	1
0.946	1	0.137	1	0.095	0	0.712	1
0.924	1	0.244	1	0.210	1	0.401	1
0.242	1	1.611	1	0.121	1	0.453	1
0.993	1	2.051	0	0.657	1	2.693	1
0.241	1	0.615	1	0.167	1	1.097	0
0.503	0	1.007	0	0.332	1	1.258	1
0.151	0	1.483	1	1.950	0	0.392	1
0.089	1	0.605	1	0.569	1	0.050	1
1.163	0	1.060	1	0.417	1	0.303	1
1.074	0	0.708	0	0.736	0	0.084	1
0.543	1	0.333	0	0.072	0	0.257	1
0.183	1	1.261	1	2.792	1	1.096	0
0.373	1	0.815	1	1.358	1	0.443	1
0.848	1	0.224	0	1.609	0	3.616	1
1.272	1	0.455	1	0.695	0	0.139	1
0.219	1	0.604	1	1.121	1	0.253	1
0.985	0	1.457	1	0.094	1	0.156	0
0.745	1	0.975	0	0.903	1	0.647	1
2.783	1	0.791	1	1.054	1	0.167	0

B=300 bootstrap samples as described in Section 4 for p=.10, .50, .75. For the chosen value $\hat{h}(p)$, the estimate $Q_n(p)$ was calculated and the bias, the standard error, and the two approximate 95% confidence intervals described above were obtained from B=1000 bootstrap samples. The results are given in Table 12. Note that the intervals from (5.3) are shifted slightly from those given by (5.1), indicating the skewness of the distribution of $Q_n(p)$.

TABLE 12. Computation Results for Simulated Data

p	$Q^o(p)$	$\hat{h}(p)$	$Q_n(p)$	Interval (5.1)	Interval (5.3)
0.10	0.1054	0.17	0.1462	(0.0849,0.2076)	(0.0953,0.2194)
0.50	0.6931	0.23	0.7440	(0.5781,0.9099)	(0.5903,0.9118)
0.75	1.3863	0.49	1.2605	(0.9466,1.5744)	(0.9735,1.5862)

The performance of the bootstrap percentile interval (5.3) depends on how well the bootstrap distribution of $Q_n^*(p)$ approximates the distribution of $Q_n(p)$. Asymptotic validity of the bootstrapped estimator $Q_n^*(p)$ can be established.

First, note that

$$|Q_n^*(p) - Q^o(p)| \le |Q_n^*(p) - Q_n(p)| + |Q_n(p) - Q^o(p)|.$$

For the first term on the right-hand-side, write

$$\begin{aligned} Q_n^*(p) - Q_n(p) &= \int_0^1 [\hat{Q}_n^*(t)-\hat{Q}_n(t)]h^{-1}(K((t-p)/h)dt \\ &= n^{-1/2}\int_0^1 [q_n^*(t)-q_n^*(p)]h^{-1}K((t-p)/h)dt \\ &\quad + n^{-1/2}\int_0^1 [q_n(p)-q_n(t)]h^{-1}K((t-p)/h)dt \\ &\quad + [\hat{Q}_n^*(p)-\hat{Q}_n(p)]\int_0^1 h^{-1}K((t-p)/h)dt \\ &\equiv n^{-1/2}I_1+n^{-1/2}I_2+I_3, \end{aligned}$$

where $q_n^*(t) = n^{1/2}[\hat{Q}_n^*(t)-Q^o(t)]$ and $q_n(t) = n^{1/2}[\hat{Q}_n(t)-Q^o(t)]$ denote the bootstrapped PL quantile process and the PL quantile process, respectively. Now, by Lemma 2 of Padgett, Lio, and Yu (1986), under the conditions on h and K stated in Section 2 and if F_o is continuous with density f_o, $f_o(Q^o(p))>0$, if f_o' is continuous, and $H(T_{F_o})<1$, where $T_{F_o} = \sup\{t:F_o(t)<1\}$, $|I_2| \to 0$ in probability as

$n\to\infty$. Also, by a proof similar to that of Lemma 2 of Padgett, Lio, and Yu (1986) using the results of Horvath and Yandell (1986), it can be shown that $|I_1|\to 0$ in probability as $n\to\infty$.

Now, under the conditions of Corollary 2.2 of Horvath and Yandell (1986),

$$\begin{aligned}|\hat{Q}_n^*(p)-\hat{Q}_n(p)| &\le \left|\frac{\hat{F}_n^*(Q^o(p))-p}{f_o(Q^o(p))} - [Q^o(p)-\hat{Q}_n^*(p)]\right| \\ &+ \left|\frac{\hat{F}_n(Q^o(p))-p}{f_o(Q^o(p))} - [Q^o(p)-\hat{Q}_n(p)]\right| \\ &+ \left|\frac{\hat{F}_n(Q^o(p))-\hat{F}_n^*(Q^o(p))}{f_o(Q^o(p))}\right| \\ &= O(n^{-1/2}(\log n)^{1/2}) + O(n^{-3/4}(\log n)^{5/4})\text{a.s.}\end{aligned}$$

for each p such that $F(Q^o(p))<1$. Thus, for such p, $|Q_n^*(p)-Q(p)|\to 0$ in probability.

Finally, since $|Q_n(p)-Q^o(p)|\to 0$ in probability (see Padgett, 1986), for p so that $F(Q^o(p))<1$, $|Q_n^*(p)-Q^o(p)|\to 0$ in probability.

Thus, the bootstrapped percentiles converge to the value $Q^o(p)$, providing large sample justification of the percentile interval. It should be noted that the bootstrap convergence results presented here refer to the theoretical bootstrap distribution of $Q_n^*(p)$ (when $B=\infty$), which in practice is estimated by Monte Carlo methods (with B=1000) as described earlier.

ACKNOWLEDGEMENT

The research of the first author was supported by the U.S. Air Force Office of Scientific Research under grant number AFOSR-84-0156 and the U.S. Army Research Office under grant number MIPR ARO 139-85.

REFERENCES

Cheng, K. F. (1984), "On Almost Sure Representations for Quantiles of the Product-Limit Estimator with Applications," Sankhya, Ser. A, 46, 426-443.

Csörgő, M. (1983), Quantile Processes with Statistical Applications (CBMS-NSF Regional Conference Series in Applied Mathematics, No. 42), Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1967), "The Two-Sample Problem with Censored Data," in Proceedings of the Fifth Berkeley Symposium (Vol. 4), Berkeley, CA: University of California Press, pp. 831-853.

Efron, B. (1980), The Jackknife, the Bootstrap, and Other Resampling Plans (CBMS-NSF Regional Conference Series in Applied Mathematics, No. 38), Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1981), "Censored Data and the Bootstrap," Journal of the American Statistical Association, 76, 312-319.

Efron, B. (1986), "Better Bootstrap Confidence Intervals," Journal of the American Statistical Association (to appear).

Efron, B. and Tibshirani, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Measures of Statistical Accuracy," Statistical Science, 1, 54-75.

Horváth, L. and Yandell, B.S. (1986), "Convergence Rates for the Bootstrapped Product Limit Process," Technical Report No. 780, University of Wisconsin, Department of Statistics.

International Mathematical and Statistical Libraries, Inc. (1982), IMSL, Houston.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 53, 457-481.

Lio, Y. L. and Padgett, W. J. (1985), "Some convergence Results for Kernel-Type Quantile Estimators under Censoring," Technical Report Number 108, University of South Carolina, Department of Statistics.

Lio, Y. L., Padgett, W. J. and Yu, K. F. (1986), "On the Asymptotic Properties of a Kernel-Type Quantile Estimator from Censored Samples," Journal of Statistical Planning and Inference, (to appear).

Marron, J. S. (1986), "Will the Art of Smoothing Ever Become a Science?" Technical Report, University of North Carolina, Chapel Hill, Department of Statistics.

Michael, J. R., Schucany, W. R. and Haas, R. W. (1976), "Generating Random Variates Using Transformations with Multiple Roots," The American Statistician, 30, 88-90.

Padgett, W. J. (1986), "A Kernel-Type Estimator of a Quantile Function from Right-Censored Data," Journal of the American Statistical Association, 81, 215-222.

Parzen, E. (1979), "Nonparametric Statistical Data Modeling," Journal of the American Statistical Association, 74, 105-121.

Nair, V. N. (1984), "Confidence Bands for Survival Functions with Censored Data: A Comparative Study," Technometrics, 26, 265-275.

Sander, J. (1975), "The Weak Convergence of Quantiles of the Product-Limit Estimator," Technical Report Number 5, Stanford University, Department of Statistics.

Yang, S. S. (1985), "A Smooth Nonparametric Estimator of a Quantile Function," Journal of the American Statistical Association, 80, 1004-1011.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): Stat. Tech. Rep. No. 116 (62N05-17)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Department of Statistics; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): University of South Carolina, Columbia, SC 29208	7b. ADDRESS (City, State and ZIP Code): Directorate of Mathematical & Information Sciences, Bolling AFB, DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR, ARO; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-84-0156
8c. ADDRESS (City, State and ZIP Code): Bolling AFB, DC 20332	10. SOURCE OF FUNDING NOS.: PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A5; WORK UNIT NO.
11. TITLE (Include Security Classification): Smooth Nonparametric Quantile Estimation under Censoring: Simulations and Bootstrap Methods.	
12. PERSONAL AUTHOR(S): W. J. Padgett and L. A. Thombs	
13a. TYPE OF REPORT: Technical; 13b. TIME COVERED FROM ____ TO ____	14. DATE OF REPORT (Yr., Mo., Day): May, 1986; 15. PAGE COUNT: 24
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES: FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Nonparametric quantile estimation; Right-censoring; Percentile interval; Bootstrap; Bandwidth selection; Monte Carlo simulations.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Based on right-censored data from a lifetime distribution F_o, a smooth nonparametric estimator of the quantile function $Q^o(p)$ is given by $Q_n(p) = h^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h)dt$, where $\hat{Q}_n(p)$ denotes the product-limit quantile function. Extensive Monte Carlo simulations indicate that at fixed p this kernel-type quantile estimator has smaller mean squared error than $\hat{Q}_n(p)$ for a range of values of the bandwidth h. A method of selecting an "optimal" bandwidth (in the sense of small estimated mean squared error) based on the bootstrap is investigated yielding results consistent with the simulation study. The bootstrap is also used to obtain interval estimates for $Q^o(p)$ after determining the optimal value of h.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED XX SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Maj. Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5027; 22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Stat. Tech. Rpt. No. 116 (62[illegible]05-17)	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (*and Subtitle*) Smooth Nonparametric Quantile Estimation Under Censoring: Simulations and Bootstrap Methods		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Stat.Tech.Rep. No. 117
7. AUTHOR(s) W. J. Padgett and L. A. Thombs		8. CONTRACT OR GRANT NUMBER(s) MIPR ARO 139-85
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of South Carolina Columbia, SC 29208		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE May 1986
		13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)		15. SECURITY CLASS. (*of this report*) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

Nonparametric quantile estimation; Right-censoring; Percentile interval; Bootstrap; Bandwidth selection; Monte Carlo simulations.

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

Based on right-censored data from a lifetime distribution F_0, a smooth nonparametric estimator of the quantile function $Q^0(p)$ is given by $Q_n(p) = h^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h)dt$, where $\hat{Q}_n(p)$ denotes the product-limit quantile function. Extensive Monte Carlo simulations indicate that at fixed p this kernel-type quantile estimator has smaller mean squared error than $\hat{Q}_n(p)$ for a range of values of the bandwidth h. A method of selecting an "optimal" bandwidth (in the sense of small estimated mean squared error) based on the bootstrap is investigated yielding results consistent with the simulation study. The bootstrap is also used to obtain interval estimates for $Q^0(p)$ after determining the optimal value of h.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

END
DTIC
8-86